K-Means Clustering: Step-by-Step Calculation

Problem

1. For the given data, classify the ten points into two classes, i.e., 0 and 1, by using k-means clustering: (4,1), (2,4), (2,3), (3,6), (4,4), (9,10), (6,8), (9,5), (8,7) and (10,8).

Solution

K-means clustering is an unsupervised learning algorithm used to classify a dataset into K clusters (classes). The goal is to minimize the variance within each cluster. In this example, we will classify data points into two clusters (Cluster 0 and Cluster 1) using k-means clustering.

Dataset and Initial Centroids

Given data points:

S. No.	x_1	x_2
1	4	1
2	2	4
3	2	3
4	3	6
5	4	4
6	9	10
7	6	8
8	9	5
9	8	7
10	10	8

Initial centroids:

$$C_1 = (1,1), \quad C_2 = (2,2)$$

Iteration 1: Distance Calculation and Cluster Assignment

The Euclidean distance between a point (x_1, x_2) and a centroid (c_1, c_2) is:

$$d = \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2}$$

0.0.1 Calculating Distances and Assigning Points

S. No.	x_1	x_2	Distance to C_1	Distance to C_2	Cluster
1	4	1	$\sqrt{(4-1)^2+(1-1)^2}=3$	$\sqrt{(4-2)^2+(1-2)^2}\approx 2.24$	C_2
2	2	4	$\sqrt{(2-1)^2+(4-1)^2}\approx 3.16$	$\sqrt{(2-2)^2+(4-2)^2}=2$	C_2
3	2	3	$\sqrt{(2-1)^2+(3-1)^2}\approx 2.24$	$\sqrt{(2-2)^2+(3-2)^2}=1$	C_2
4	3	6	$\sqrt{(3-1)^2+(6-1)^2}\approx 5.39$	$\sqrt{(3-2)^2+(6-2)^2}\approx 4.12$	C_2
5	4	4	$\sqrt{(4-1)^2+(4-1)^2}\approx 4.24$	$\sqrt{(4-2)^2+(4-2)^2}\approx 2.83$	C_2
6	9	10	$\sqrt{(9-1)^2+(10-1)^2}\approx 12.04$	$\sqrt{(9-2)^2+(10-2)^2}\approx 10.63$	C_2
7	6	8	$\sqrt{(6-1)^2+(8-1)^2}\approx 8.60$	$\sqrt{(6-2)^2+(8-2)^2}\approx 7.21$	C_2
8	9	5	$\sqrt{(9-1)^2+(5-1)^2}\approx 8.94$	$\sqrt{(9-2)^2+(5-2)^2}\approx 7.62$	C_2
9	8	7	$\sqrt{(8-1)^2+(7-1)^2}\approx 9.22$	$\sqrt{(8-2)^2+(7-2)^2}\approx 7.81$	C_2
10	10	8	$\sqrt{(10-1)^2+(8-1)^2}\approx 11.40$	$\sqrt{(10-2)^2+(8-2)^2}=10$	C_2

All points are initially closer to C_2, so we need to update the centroids.

Update Centroids

Since all points are assigned to C_2, we recalculate C_2 as the mean of all points:

$$C_2=\left(\frac{4+2+2+3+4+9+6+9+8+10}{10},\frac{1+4+3+6+4+10+8+5+7+8}{10}\right)=(5.7,5.6)$$

Iteration 2: Distance Calculation and Cluster Assignment

Recalculate distances with updated centroids:

S. No.	x_1	x_2	Distance to $C_1=(1,1)$	Distance to $C_2=(5.7,5.6)$	Cluster
1	4	1	3	$\sqrt{(4-5.7)^2+(1-5.6)^2}\approx 4.81$	C_1
2	2	4	3.16	$\sqrt{(2-5.7)^2+(4-5.6)^2}\approx 4.03$	C_1
3	2	3	2.24	$\sqrt{(2-5.7)^2+(3-5.6)^2}\approx 4.51$	C_1
4	3	6	5.39	$\sqrt{(3-5.7)^2+(6-5.6)^2}\approx 2.82$	C_2
5	4	4	4.24	$\sqrt{(4-5.7)^2+(4-5.6)^2}\approx 2.40$	C_2
6	9	10	12.04	$\sqrt{(9-5.7)^2+(10-5.6)^2}\approx 5.59$	C_2
7	6	8	8.60	$\sqrt{(6-5.7)^2+(8-5.6)^2}\approx 2.41$	C_2
8	9	5	8.94	$\sqrt{(9-5.7)^2+(5-5.6)^2}\approx 3.35$	C_2
9	8	7	9.22	$\sqrt{(8-5.7)^2+(7-5.6)^2}\approx 2.59$	C_2
10	10	8	11.40	$\sqrt{(10-5.7)^2+(8-5.6)^2}\approx 4.75$	C_2

Updated clusters: Points $1,2,3$ assigned to C_1 and points $4,5,6,7,8,9,10$ assigned to C_2.

0.0.2 Update Centroids Again

$$C_1 = \left(\frac{4+2+2}{3}, \frac{1+4+3}{3}\right) = (2.67, 2.67)$$

$$C_2 = \left(\frac{3+4+9+6+9+8+10}{7}, \frac{6+4+10+8+5+7+8}{7}\right) = (7, 6.86)$$

Iteration 3: Distance Calculation and Cluster Assignment

Recalculate distances with updated centroids:

S. No.	x_1	x_2	Distance to $C_1 = (2.67, 2.67)$	Distance to $C_2 = (7, 6.86)$	Cluster
1	4	1	$\sqrt{(4-2.67)^2 + (1-2.67)^2} \approx 2.13$	$\sqrt{(4-7)^2 + (1-6.86)^2} \approx 6.60$	C_1
2	2	4	$\sqrt{(2-2.67)^2 + (4-2.67)^2} \approx 1.50$	$\sqrt{(2-7)^2 + (4-6.86)^2} \approx 5.78$	C_1
3	2	3	$\sqrt{(2-2.67)^2 + (3-2.67)^2} \approx 0.75$	$\sqrt{(2-7)^2 + (3-6.86)^2} \approx 6.31$	C_1
4	3	6	$\sqrt{(3-2.67)^2 + (6-2.67)^2} \approx 3.35$	$\sqrt{(3-7)^2 + (6-6.86)^2} \approx 4.09$	C_1
5	4	4	$\sqrt{(4-2.67)^2 + (4-2.67)^2} \approx 1.88$	$\sqrt{(4-7)^2 + (4-6.86)^2} \approx 4.13$	C_1
6	9	10	$\sqrt{(9-2.67)^2 + (10-2.67)^2} \approx 9.67$	$\sqrt{(9-7)^2 + (10-6.86)^2} \approx 3.73$	C_2
7	6	8	$\sqrt{(6-2.67)^2 + (8-2.67)^2} \approx 6.25$	$\sqrt{(6-7)^2 + (8-6.86)^2} \approx 1.50$	C_2
8	9	5	$\sqrt{(9-2.67)^2 + (5-2.67)^2} \approx 6.76$	$\sqrt{(9-7)^2 + (5-6.86)^2} \approx 2.74$	C_2
9	8	7	$\sqrt{(8-2.67)^2 + (7-2.67)^2} \approx 6.78$	$\sqrt{(8-7)^2 + (7-6.86)^2} \approx 1.01$	C_2
10	10	8	$\sqrt{(10-2.67)^2 + (8-2.67)^2} \approx 9.04$	$\sqrt{(10-7)^2 + (8-6.86)^2} \approx 3.18$	C_2

0.0.3 Update Centroids for the Fourth Iteration

$$C_1 = \left(\frac{4+2+2+3+4}{5}, \frac{1+4+3+6+4}{5}\right) = (3, 3.6)$$

$$C_2 = \left(\frac{9+6+9+8+10}{5}, \frac{10+8+5+7+8}{5}\right) = (8.4, 7.6)$$

Iteration 4: Distance Calculation and Cluster Assignment

Recalculate distances with updated centroids:

S. No.	x_1	x_2	Distance to $C_1 = (3, 3.6)$	Distance to $C_2 = (8.4, 7.6)$	Cluster
1	4	1	$\sqrt{(4-3)^2+(1-3.6)^2} \approx 2.83$	$\sqrt{(4-8.4)^2+(1-7.6)^2} \approx 7.57$	C_1
2	2	4	$\sqrt{(2-3)^2+(4-3.6)^2} \approx 1.08$	$\sqrt{(2-8.4)^2+(4-7.6)^2} \approx 7.56$	C_1
3	2	3	$\sqrt{(2-3)^2+(3-3.6)^2} \approx 1.17$	$\sqrt{(2-8.4)^2+(3-7.6)^2} \approx 7.84$	C_1
4	3	6	$\sqrt{(3-3)^2+(6-3.6)^2} = 2.4$	$\sqrt{(3-8.4)^2+(6-7.6)^2} \approx 5.59$	C_1
5	4	4	$\sqrt{(4-3)^2+(4-3.6)^2} \approx 1.08$	$\sqrt{(4-8.4)^2+(4-7.6)^2} \approx 5.52$	C_1
6	9	10	$\sqrt{(9-3)^2+(10-3.6)^2} \approx 8.61$	$\sqrt{(9-8.4)^2+(10-7.6)^2} \approx 2.47$	C_2
7	6	8	$\sqrt{(6-3)^2+(8-3.6)^2} \approx 5.02$	$\sqrt{(6-8.4)^2+(8-7.6)^2} \approx 2.47$	C_2
8	9	5	$\sqrt{(9-3)^2+(5-3.6)^2} \approx 6.17$	$\sqrt{(9-8.4)^2+(5-7.6)^2} \approx 2.67$	C_2
9	8	7	$\sqrt{(8-3)^2+(7-3.6)^2} \approx 6.04$	$\sqrt{(8-8.4)^2+(7-7.6)^2} \approx 0.72$	C_2
10	10	8	$\sqrt{(10-3)^2+(8-3.6)^2} \approx 8.17$	$\sqrt{(10-8.4)^2+(8-7.6)^2} \approx 1.63$	C_2

Clusters remain unchanged after the fourth iteration.

Conclusion

The k-means clustering algorithm converged after four iterations. The final clusters are:

- **Cluster 1 (Class 0):** Points 1, 2, 3, 4, 5
- **Cluster 2 (Class 1):** Points 6, 7, 8, 9, 10